

Estimation of Discretized Motion of Pedestrians by the Decision-Making Model

Pavel Hrabák and Ondřej Ticháček and Vladimíra Sečkárová

Abstract The contribution gives a micro-structural insight into the pedestrian decision process during an egress situation. A method how to extract the decisions of pedestrians from the trajectories recorded during the experiments is introduced. The underlying Markov decision process is estimated using the finite mixture approximation. Furthermore, the results of this estimation can be used as an input to the optimization of a Markov decision process for one ‘clever’ agent. This agent optimizes his strategy of motion with respect to different reward functions, minimizing the time spent in the room or minimizing the amount of inhaled CO.

1 Introduction

This study can be used as an auxiliary calibration tool for microscopic models of pedestrian flow with spatially discretized motion of agents, as e.g. floor-field model [4] or optimal-steps model [7]. The results can be applied in the navigation robotic systems [8]. The introduced method builds upon the floor-field model. Thanks to the restriction to the discretized motion of pedestrian we are able to express the local decisions of pedestrian in the terms of Markov decision process [5].

For the analysis of the real data, we use the experimental data from a passing-through experiment [1]. In this experiment, pedestrians were instructed to pass through a simple room equipped by one entrance with controlled inflow and one exit of the width 60 cm. Since we are mainly interested in the pedestrian interaction, we used the data from the rear camera covering the space of 2.5 m in front of the exit and short part of the corridor behind the exit.

Throughout the article, we use the notation related to Markov decision processes (MDP) adopted from the book [5]. The main task of the contribution is to express

P. Hrabák, O. Ticháček and V. Sečkárová
Institute of Information Theory and Automation, Czech Academy of Sciences, Pod Vodarenskou
vezi 4, CZ-182 08, Prague 8, Czech Republic. e-mail: hrabak@utia.cas.cz

the basic entries of the MDP theory in the scope of pedestrian flow dynamics. This is necessary to use the optimization technique described in [5, Chapter 4].

2 Basic Concept

Let us describe the MDP in general. The considered decision process (DP) is characterized by a sequence $(s_1, a_1, s_2, a_2, \dots, s_{T-1}, a_{T-1}, s_T)$ of states $s_t \in S$ and performed actions $a_t \in A$. Here T plays the role of a finite time horizon used for the optimization. At time t an agent, who is making the decision, observes the system to be in state s_t and based on this observation performs an action a_t with conditional probability $p_t(a_t | s_t)$. The system reacts to the action stochastically and the state changes to s_{t+1} with conditional probability $p_t(s_{t+1} | s_t, a_t)$. This probability can be understood as the agent's image of the environment behaviour. The Markov property is hidden in the fact that both, the decision part $p_t(a_t | s_t)$ and the environmental model $p_t(s_{t+1} | s_t, a_t)$, depend only on the situation at time t . Then the probability of a sequence $(s_1, a_1, s_2, a_2, \dots, s_{T-1}, a_{T-1}, s_T)$ is given as

$$\Pr(s_1, a_1, s_2, a_2, \dots, s_{T-1}, a_{T-1}, s_T) = p(s_1) \prod_{t=1}^{T-1} p_t(a_t | s_t) p_t(s_{t+1} | s_t, a_t) \quad (1)$$

This concept can be easily applied to the floor-field model [4] (For more details about the model we refer the reader to the book [6]). The floor-field model is a particle hopping model defined on a rectangular lattice L representing the discretized layout of the simulated facility. Particles are hopping between cells stochastically according to the hopping probabilities, which are influenced by the static floor field S . Usually $S(y) = \text{dist}(y, \mathbf{E})$ refers to the distance of the cell y to the exit \mathbf{E} in defined metric dist . Let the state of the system at time t be denoted by $\tau_t \in \{0, 1\}^L$, where $\tau(x) = 1$ refers to an occupied cell and $\tau(x) = 0$ to an empty cell. Let further $n_t = \sum_x \tau_t(x)$ be the number of agents in the lattice at time t .

In each algorithm step $t \rightarrow t + 1$, every agent $i \in \{1, \dots, n_t\}$ chooses his future position $y_{i,t}$ given he is sitting in $x_{i,t}$ with probability

$$p(y_{i,t} | x_{i,t}, \tau_t) \propto \exp\{-S(y_{i,t})\} \mathbf{1}_{\{\text{dist}(x_{i,t}, y_{i,t}) \leq 1\}} \quad (2)$$

according to floor-field model. The model of the environment is then a consequence of the choices of future positions of all agents, i.e., the dynamics is driven by the environment model

$$p(\tau_{t+1} | \tau_t, y_{i,t}, i \in \{0, \dots, n_t\}) = F(p(y_{i,t} | x_{i,t}, \tau_t), i \in \{0, \dots, n_t\}), \quad (3)$$

where F is a function that reflects conflicting situation where two agents choose the same target cell¹.

¹ Without the conflicting situations the function is just a product of the entries.

3 Estimating $p_t(a_t | s_t)$

This section focuses on the probability probabilistic decision $p_t(a_t | s_t)$ from the data recorded during evacuation experiment [1].

Let us assume that the pedestrians act similarly to the floor-field particles, i.e., all pedestrians are following the same decision strategy, which does not change in time and space. Furthermore, we assume that pedestrians react only on their immediate neighbourhood reflecting the direction towards the exit, but not their absolute position. Therefore, the state s_t in MDP can be associated with the state of the immediate neighbourhood. Contrary to the floor-field model we consider the neighbourhood to be oriented with respect to the direction towards the exit. The actions are associated with direction angle a pedestrian can choose, see Figure 1.

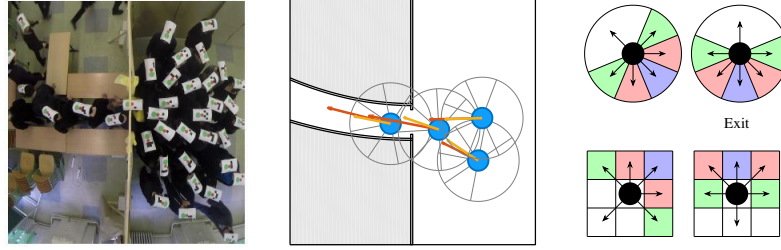


Fig. 1 Transformation of trajectory record to actions. Trajectories are extracted from the video records (left, taken from [1]), transformed to motion within sectors (middle) with respect to the direction towards exit, and interpreted as motion in the lattice (right).

The experimental data for trajectory analyses have been provided by our colleague Marek Bukáček (Czech Technical University). The data are in the form of paths records $(\mathbf{x}_i(t), t \in [t_i^{\text{in}}, t_i^{\text{out}}])$, where t_i^{in} and t_i^{out} is the time of the first and the last appearance of the pedestrian i on the screen respectively. $\mathbf{x}_i(t)$ is the position of the pedestrian on the screen at time t . To match the discrete nature of the decision-making process, the motion of pedestrians has been discretized in time with the discretization step $\Delta t = 1$ s. The vector of motion at time t is then $\Delta \mathbf{x}_i(t) = \mathbf{x}_i(t + \Delta t) - \mathbf{x}_i(t)$ and the direction of motion $\vartheta_i(t)$ is an angle given by

$$\cos \vartheta_i(t) = \frac{[\mathbf{E} - \mathbf{x}_i(t)] \cdot \Delta \mathbf{x}_i(t)}{\|\mathbf{E} - \mathbf{x}_i(t)\| \cdot \|\Delta \mathbf{x}_i(t)\|} \quad (4)$$

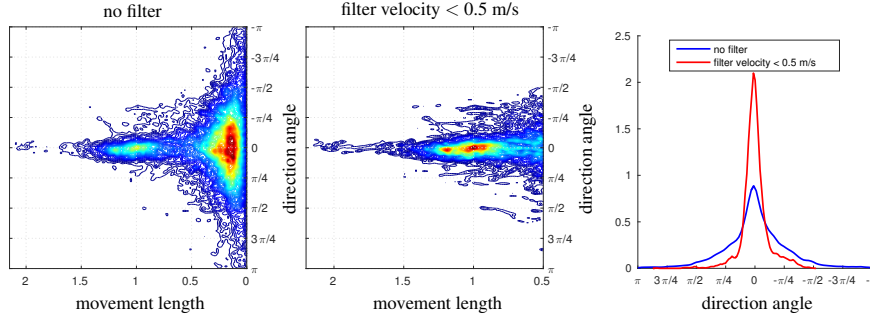
This angle is then associated to the action $a \in A$ according to the Table 1. The set of actions is $A = \{\otimes, \leftarrow, \nearrow, \swarrow, \uparrow, \downarrow, \rightarrow\}$, e.g., an angle $\vartheta_i(t) = 20^\circ$ corresponds to the forward motion \leftarrow , while $\vartheta_i(t) = 30^\circ$ corresponds to the left-forward motion \swarrow . Every motion performed with velocity $v_i(t) = \|\Delta \mathbf{x}_i(t)\|/\Delta t$ less than 0.5 m/s has been considered as standing (\otimes).

In Figure 2 the frequency of chosen direction with respect to the direction angle is plotted. Two main clusters for forward stepping and standing can be distinguished,

Table 1 Action set in detail. Direction angle towards the exit is 0. Colours refer to Figure 1.

Action:	\otimes	\leftarrow	\nwarrow	\swarrow	\uparrow	\downarrow	\rightarrow
Angle:	\emptyset	$(-\frac{\pi}{8}, \frac{\pi}{8})$	$(-\frac{\pi}{8}, -\frac{3\pi}{8})$	$(\frac{3\pi}{8}, \frac{\pi}{8})$	$(-\frac{3\pi}{8}, -\frac{5\pi}{8})$	$(\frac{5\pi}{8}, \frac{3\pi}{8})$	$(-\pi, -\frac{5\pi}{8}) \cup (\frac{5\pi}{8}, \pi)$
Color:	black	blue	red	red	green	green	white

the latter dominates. The data are aggregated over all pedestrians and all records for each. From this graph we can conclude that the majority of pedestrians preferred standing in line and moving forward centimetre by centimetre rather than trying to push through the crowd or overrunning it.

**Fig. 2** The frequency of occurrence of the motion length in given direction angle. The graphs are oriented similarly to the snapshot from the experiment in Figure 1. Left: without the filter on standing. Middle: after filtering out the motion with velocity less then 0.5 m/s. Right: Gaussian-kernel estimation of the distribution of chosen direction.

For the estimation of the decision process we associate the state s_t in the decision-process $p(a_t | s_t)$ with the occupancy of the immediate neighbourhood. By the neighbourhood $N(\mathbf{x})$ of a position \mathbf{x} we understand a circle around \mathbf{x} with the radius 0.75 m (maximal step size) divided into 6 sectors $\{\leftarrow, \nwarrow, \swarrow, \uparrow, \downarrow, \rightarrow\}$ defined in previous section. The state $s_{i,t}$ for the decision $p(a_{i,t} | s_{i,t})$ of the agent i is then a vector from $\{0, 1\}^6$, where $s_{i,t}(y) = 0$ for empty sector y and $s_{i,t}(y) = 1$ for occupied sector. Here $y \in \{\leftarrow, \nwarrow, \swarrow, \uparrow, \downarrow, \rightarrow\}$. The sector is considered occupied if it contains at least one position vector of another agent or if it is covered by a wall by at least 40 %. Since the data are aggregated over all pedestrians, the index i will be further omitted.

Most natural way how to estimate the decision process $p(a_t | s_t)$ is to compare the frequency of chosen directions (actions) given the occupancy of the neighbourhood. However, this method fails in the case of the trajectory data from considered experiment, since most of the combinations (a_t, s_t) appear very rarely. For this reason we applied the approximation of the decision by finite mixture model with forgetting [3]. The idea consists in approximation of the complex decision process $p(a_t | s_t)$ by the convex combination of marginal decision processes $\Theta(a_t | s_t(y))$,

i.e.,

$$p(a_t | s_t) \approx p(a_t | \Theta, s_t) = \sum_{y \in N(\mathbf{x}_t)} \alpha_y \Theta(a_t | s_t(y)), \quad (5)$$

where α_y is the coefficient of influence of the state of y to the decision; $\sum_y \alpha_y = 1$. For more details see [2].

The resulting values of the mixture model are given in Table 2. The following phenomena can be observed analysing the values in the table. The occupancy of a neighbouring sector almost always contributes with the highest value to the decision “to stand” (\otimes). However, a free neighbouring sector does not always tend to imply motion, see the forward sector (\leftarrow). The most diverse influence of the empty and occupied states show the “slightly right” (\nearrow) and “slightly left” (\swarrow) sectors. The explanation for this may be a zipper-like effect of agents passing through a narrow exit and corridor.

The table also implies, that the occupancy of right and left sectors (\uparrow, \downarrow) does not play a significant role in agent’s decision as it does not restrain him from moving in desired (forward) direction. Finally, although in principle, the occupancy of the back sector (\rightarrow) should not affect the agent’s decision in his desire to go straight, this sector is mostly occupied if the agent is in a high-density situation (eg. a jam) and therefore it’s occupancy reflects the agent’s (in)ability to move at all.

Table 2 Influence of regressors y to decision a for two states of sector occupancy (\circ symbolizes empty sector, \bullet occupied). $\hat{\alpha}$ is the weight of the regressor.

y	\leftarrow		\nearrow		\swarrow		\uparrow		\downarrow		\rightarrow	
$\hat{\alpha}$	0.9212		0.0749		0.0014		0.0002		0.0002		0.0021	
a	\circ	\bullet	\circ	\bullet	\circ	\bullet	\circ	\bullet	\circ	\bullet	\circ	\bullet
\otimes	0.76	0.94	0.03	0.97	0.05	0.52	0.11	0.12	0.12	0.11	0.07	0.69
\leftarrow	0.22	0.06	0.63	0.01	0.73	0.12	0.39	0.23	0.35	0.24	0.68	0.10
\nearrow	0.01	0.01	0.02	0.01	0.05	0.06	0.09	0.11	0.10	0.11	0.06	0.04
\swarrow	0.01	0.01	0.27	0.00	0.04	0.08	0.09	0.11	0.10	0.11	0.06	0.04
\uparrow	0.00	0.00	0.02	0.00	0.04	0.06	0.08	0.11	0.09	0.11	0.04	0.04
\downarrow	0.00	0.00	0.02	0.00	0.03	0.05	0.08	0.11	0.08	0.11	0.04	0.03
\rightarrow	0.00	0.00	0.02	0.00	0.06	0.10	0.15	0.20	0.16	0.22	0.06	0.06

4 Optimizing $p_t(a_t | s_t)$

This chapter offers an alternative view on the application of MDP to pedestrian flow modelling. Let the result of previous section be used as the behavioural frame of the majority of pedestrians, which determines the environmental model $p_t(s_{t+1} | s_t, a_t)$ described by equation (3). The aim of this section is to equip one of the pedestrian agent by an optimal decision strategy how to move among the pedestrians following the majority behaviour.

Let us in the following, for simplicity, return back to the Floor-field basis of the simulation. Consider that there is one “clever” particle among the ordinary indistinguishable floor-field particles behaving according to (2) or (5). By the optimal strategy of the clever particle we understand the sequence $(d_1, d_2, \dots, d_{T-1})$, where $d_t(s_t) = q$ is the distribution on the set of actions A playing the role of the decision process, i.e., $p(a_t | s_t) = q(a_t)$. The strategy is optimized with respect to given reward function $R(s_1, a_1, \dots, s_T)$, which can be used to model different preferences of the clever particle.

Contrarily to Section 3 we consider in the following the position and the orientation to be absolute, i.e., there are not preferred positions on the lattice regarding the distance or orientation towards the lattice. The optimizing algorithm is supposed to find the shortest path given the maximal reward R itself. The state of the system is expressed by the position of the clever agent x_t and by the set of positions of all floor-field particles z_t , where $|z_t| = n_t - 1$ (compare to previous section, where only the neighbouring pedestrians play role). Let the positions be numbered by natural numbers as shown by an example in Figure 3.

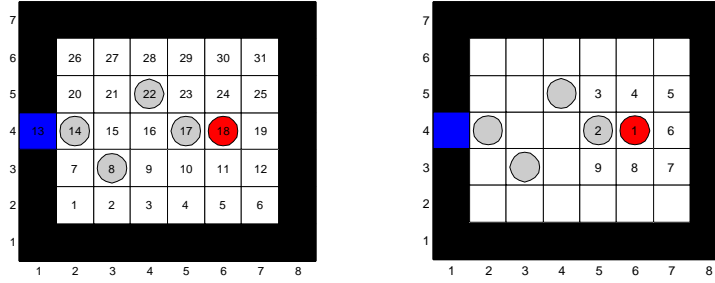


Fig. 3 An example of the state of the lattice with one clever agent (red) and 4 floor-field particles (gray). Left: the state is $s_t = (x_t, z_t)$, where $x_t = 18$, $z_t = \{8, 14, 17, 22\}$. Right: the numbering of directions corresponding to actions $a_t \in \{1, \dots, 9\}$.

The actions an agent can choose are related to the 8 neighbouring sites and a possibility to stay in current position. Therefore the action set can be chosen as $A = \{1, 2, \dots, 9\}$, where the directions are numbered as depicted in Figure 3, i.e., $a_t = 1$ means to choose as next target site the current position $x_t = 18$; $a_t = 2$ corresponds to the target site $x_t + (-1, 0) = 17$, etc. Here we note that, similarly to the floor-field, the chosen target site can be entered by a near floor-field particle. The choice of the target site is therefore influenced by the probability that other particles can change their positions.

For the purposes of this contribution we have chosen a simple updating scheme in which the clever agent moves after all other particles performed their actions. Then the environmental model $p_t(s_{t+1} | s_t, a_t) = p(x_{t+1}, z_{t+1} | x_t, z_t, a_t)$ decomposes into a stochastic part $p(z_{t+1} | x_t, z_t)$ and a deterministic part $p(x_{t+1} | x_t, a_t, z_{t+1})$ as

$$p(x_{t+1}, z_{t+1} | x_t, z_t, a_t) = p(z_{t+1} | x_t, z_t) p(x_{t+1} | x_t, a_t, z_{t+1}). \quad (6)$$

In our case, the transition probability (5) can be simplified to the form

$$p(z_{t+1} | x_t, z_t) \propto \exp\{-U(z_{t+1})\}, \quad U(z_{t+1}) = \sum_{j=1}^{|z_{t+1}|} \text{dist}(z_{t+1}(j), \mathbf{E}) \quad (7)$$

for all states z_{t+1} reachable from z_t by the motion of floor-field particles by one site. The introduced potential U supports the states in which particles are closer to the exit and therefore suppresses the random motion away from the exit.

The above mentioned concept fits the finite time optimization of the MDP strategies using the backward induction algorithm described in [5, Section 4.5]. The final step is to define properly the reward function R . The reward function is defined as a cumulation of local rewards $r_t(s_t, a_t)$ and the final reward $v_T(s_T)$, i.e.,

$$R = \sum_{t=1}^{T-1} r_t(s_t, a_t) + v_T(s_T), \quad r_t(s_t, a_t) = \sum_{s_{t+1}} p(s_{t+1} | s_t, a_t) r_t(s_{t+1}, s_t, a_t). \quad (8)$$

The final reward is the same for all agents preferences taking into account the distance to the exit multiplied by a factor of 2, i.e., $v_T(s_T) = -2 \text{dist}(x_T, \mathbf{E})$. The local reward then reflects the agent's preferences. We introduce two main approaches: minimizing the time spent in the room and minimizing the amount of inhaled carbon monoxide (CO) related to the aim to minimize number of lost conflicts.

The reward function minimizing the time simply subtracts one reward unit for each step an agent spends outside the exit, i.e.,

$$r_t(s_{t+1}, s_t, a_t) = \begin{cases} 0 & x = \mathbf{E}, \\ -1 & x \neq \mathbf{E}. \end{cases} \quad (9)$$

The reward function minimizing the amount of inhaled CO takes into account the possibility that the agent can choose a site which becomes occupied by another particle. Such choice can be interpreted as running to another pedestrian, which causes a significant loss of energy with no improvement of the distance to the exit. Such situation costs 2 reward units, while standing only one half. Therefore

$$r_t(s_{t+1}, s_t, a_t) = \begin{cases} 0 & x_t = \mathbf{E}, \\ -1/2 & a_t = 1, \\ -1 & a_t \neq 1, x_{t+1} \notin z_{t+1}, \\ -2 & a_t \neq 1, x_{t+1} \in z_{t+1}. \end{cases} \quad (10)$$

5 Conclusions and Future Plans

The main goal of this paper was to introduce a concept of Markov decision process (MDP) to the pedestrian flow simulation. Two aspects have been studied by means

of this concept: the estimation of pedestrian behaviour within crowded area and the optimization of the decision with respect to given pedestrian preferences. Both approaches are motivated by the cellular floor-field model used for simulation of pedestrian evacuation.

The estimation of pedestrian behaviour have been analysed from experimental trajectories. By means of the space discretization and finite mixture approximation we have been able to extract the pedestrians decision in relation to the occupation of his immediate neighbourhood. The analysis showed that the main influence to the decision has the occupation of the area in the forward direction towards the exit. Further, most of the decisions pedestrians performed was to move forward or stay at the position. The over-running of the crowd was rather a rare event.

The results of the experiment analyses can be then used as the input to the optimization task of one ‘clever’ agent among usual floor-field particles. We have introduced a technique of expressing the pedestrian evacuation model in terms of the MDP. Furthermore, two different reward functions have been introduced to simulate different preferences of the clever agent: to minimize the time spent in the room and to minimize the amount of inhaled carbon oxygen, i.e. minimizing the number of conflicts. In the future we plan to test the combination of those two strategies in order to prove that optimal is the combination of the two above mentioned strategies.

Acknowledgements This work was supported by the Czech Science Foundation under the grant GA13-13502S. We want to thank our colleague Marek Bukáček from Czech Technical University for the provided experimental data.

References

1. Bukáček, M., Hrabák, P., Krbálek, M.: Experimental study of phase transition in pedestrian flow. In: W. Daamen, D.C. Duives, S.P. Hoogendoorn (eds.) *Pedestrian and Evacuation Dynamics 2014, Transportation Research Procedia*, vol. 2, pp. 105–113. Elsevier Science B.V. (2014)
2. Hrabák, P., Ticháček, O.: Prediction of pedestrian decisions during the egress situation: Application of recursive estimation of high-order Markov chains using the approximation by finite mixtures. Tech. Rep. 2346, ÚTIA AV ČR, POBox 18, 182 08 Prague 8, Czech Republic (2015)
3. Kárný, M.: Recursive estimation of high-order markov chains: Approximation by finite mixtures. *Information Sciences* **326**, 188–201 (2016)
4. Kirchner, A., Schadschneider, A.: Simulation of evacuation processes using a bionics-inspired cellular automaton model for pedestrian dynamics. *Physica A: Statistical Mechanics and its Applications* **312**(12), 260 – 276 (2002)
5. Puterman, M.L.: *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, 1st edn. John Wiley & Sons, Inc., New York, NY, USA (1994)
6. Schadschneider, A., Chowdhury, D., Nishinari, K.: *Stochastic Transport in Complex Systems: From Molecules to Vehicles*. Elsevier Science B. V., Amsterdam (2010)
7. Seitz, M.J., Köster, G.: Natural discretization of pedestrian movement in continuous space. *Physical Review E* **86**, 046,108 (2012)
8. Spaan, M., Veiga, T., Lima, P.: Decision-theoretic planning under uncertainty with information rewards for active cooperative perception. *Autonomous Agents and Multi-Agent Systems* **29**(6), 1157–1185 (2015)